Iniziato	martedì, 7 gennaio 2020, 15:04
Stato	Completato
Terminato	martedì, 7 gennaio 2020, 15:32
Tempo impiegato	28 min. 15 secondi
Punteggio	13,00/15,00
Valutazione	26,00 su un massimo di 30,00 (87%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In data preprocessing, which of the following are the objectives of the *aggregation* of attributes

Scegli una o più alternative:

- [x] **a. Obtain a less detailed scale** ✔
- [ ] **b. Obtain a more detailed description of data**
- [x] **c. Reduce the variability of data** ✔
- [x] **d. Reduce the number of attributes or distinct values** ✔

Risposta corretta.

Le risposte corrette sono: Reduce the number of attributes or distinct values, Obtain a less detailed scale, Reduce the variability of data

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the effect of the *curse of dimensionality*

Scegli un'alternativa:

- [ ] **a. When the number of dimensions increases the computing power necessary to compute the distances becomes too high**
- [ ] **b. When the number of dimensions increases the results tend to be prone to overfitting**
- [x] **c. When the number of dimensions increases the euclidean distance becomes less effective to discriminate between points in the space** ✔
- [ ] **d. When the number of dimensions increases the classifiers cannot be correctly tuned**

Risposta corretta.

La risposta corretta è: When the number of dimensions increases the euclidean distance becomes less effective to discriminate between points in the space

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (One or more)

Scegli una o più alternative:

- **a. K-means is very sensitive to the initial assignment of the centers ✔ No, being based on distances, if the number of attributes is very large k-means is prone to the *curse of dimensionality***
- **b. K-means always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters.**
- **c. K-means is quite efficient even for large datasets ✔ No, k-means finds a local minimum of the distortion for an assigned number of clusters**
- **d. Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters. ✔**

Your answer is correct.

Le risposte corrette sono: Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters., K-means is quite efficient even for large datasets, K-means is very sensitive to the initial assignment of the centers

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Consider the transactional dataset below

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *support* of the rule A,C ⇒ B?

Scegli un'alternativa:

- **a. 40%**
- **b. 50%**
- **c. 20% ✔ 1 / 5**
- **d. 100%**

Risposta corretta.

La risposta corretta è: 20%

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What does K-means try to minimise?

Scegli un'alternativa:

- a. The *separation*, that is the sum of the squared distances of each point with respect to its centroid
- b. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid ✔
- c. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters
- d. The *separation*, that is the sum of the squared distances of each cluster centroid with respect tho the global centroid of the dataset

Risposta corretta.

La risposta corretta è: The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

Domanda **6**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

Which is the main purpose of *smoothing* in Bayesian classification?

Scegli un'alternativa:

- a. Dealing with missing values ✖
- b. Classifying an object containing attribute values which are missing from some classes in the training set
- c. Reduce the variability of the data
- d. Classifying an object containing attribute values which are missing from some classes in the test set

Risposta errata.

La risposta corretta è: Classifying an object containing attribute values which are missing from some classes in the training set

Domanda **7**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

In a decision tree, the number of objects in a node...

Scegli un'alternativa:

- a. ...is smaller than the number of objects in its ancestor
- b. ...is not related to the number of objects in its ancestor
- c. ...is bigger than the number of objects in its ancestor
- d. ...is smaller than or equal to the number of objects in its ancestor ✖ It cannot be equal, if you split the decrease in size is at least one

Risposta errata.

La risposta corretta è: ...is smaller than the number of objects in its ancestor

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following measure can be used as an alternative to the *Information Gain?*

Scegli un'alternativa:

- a. Silhouette Index
- b. Rand Index
- c. Jaccard Index
- d. Gini Index ✔

Your answer is correct.

La risposta corretta è: Gini Index

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Why do we *prune* a decision tree?

Scegli un'alternativa:

- a. To eliminate parts of the tree where the decision could generate *underfitting*
- b. To eliminate attributes which could be influenced by random effects
- c. To eliminate rows of the dataset which could be influenced by random effects
- d. To eliminate parts of the tree where the decisions could be influenced by random effects ✔

Your answer is correct.

La risposta corretta è: To eliminate parts of the tree where the decisions could be influenced by random effects

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following *is not* an objective of feature selection

Scegli un'alternativa:

- a. Reduce the effect of noise
- b. Avoid the *curse of dimensionality*
- c. Select the features with higher range, which have more influence on the computations ✔
- d. Reduce time and memory complexity of the mining algorithms

Risposta corretta.

La risposta corretta è: Select the features with higher range, which have more influence on the computations

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following clustering methods is *not* based on distances between objects?

Scegli un'alternativa:

- **a. Expectation Maximization** ✔
- **b. Hierarchical Agglomerative**
- **c. DBSCAN**
- **d. K-Means**

Your answer is correct.

La risposta corretta è: Expectation Maximization

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following characteristic of data can reduce the effectiveness of DBSCAN?

Scegli un'alternativa:

- **a. Presence of outliers**
- **b. Clusters have concavities**
- **c. Presence of clusters with different densities** ✔
- **d. All the variables are the same range of values**

Your answer is correct.

La risposta corretta è: Presence of clusters with different densities

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following types of data allows the use of the euclidean distance?

Scegli un'alternativa:

- **a. Ordered data**
- **b. Document representations**
- **c. Transactional data**
- **d. Points in a vector space** ✔

Your answer is correct.

La risposta corretta è: Points in a vector space

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- **TP = True Positives**
- **TN = True Negatives**
- **FP = False Positives**
- **FN = False Negatives**

which of the formulas below computes the *precision* of a binary classifier?

Scegli un'alternativa:

- **a. TP / (TP + FN)**
- **b. TP / (TP + FP)** ✔ **This is also called *positive predictive value*, which is the number of detected true positives divided by the total number of elements predicted as positive**
- **c. TN / (TN + FP)**
- **d. (TP + TN) / (TP + FP + TN + FN)**

Risposta corretta.

La risposta corretta è: TP / (TP + FP)

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Match the rule evaluation formulas with their names

$$\frac{conf(A \Rightarrow C)}{sup(C)}$$ — **Lift** ✔

$$\frac{sup(A \Rightarrow C)}{sup(A)}$$ — **Confidence** ✔

$$\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$$ — **Conviction** ✔

$$sup(A \cup C) - sup(A)sup(C)$$ — **Leverage** ✔

Your answer is correct.

La risposta corretta è: $\frac{conf(A \Rightarrow C)}{sup(C)}$ **→ Lift,**

$\frac{sup(A \Rightarrow C)}{sup(A)}$ **→ Confidence,**

$\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$ **→ Conviction,**

$sup(A \cup C) - sup(A)sup(C)$ **→ Leverage**